The multivariate normal (MVN) distribution

This is the most important continuous joint distribution, and is often a natural choice for modelling multivariate data.

For each of these individually, we would probably choose a (univariate) normal distribution as our model, and the multivariate normal distribution provides a way of modelling the way that they vary together where each of the marginal distributions is univariate normal.

The independent bivariate case

Consider two independent random variables U and V each following respectively (univariate) normal distributions.

Say $U \sim N(\mu_1, \sigma_1^2)$ and $V \sim N(\mu_2, \sigma_2^2)$ with p.d.f.'s

$$f_U(u) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{ -\frac{(u-\mu_1)^2}{2\sigma_1^2} \right\}$$

and

$$f_V(v) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{ -\frac{(v-\mu_2)^2}{2\sigma_2^2} \right\}$$

Joint p.d.f.

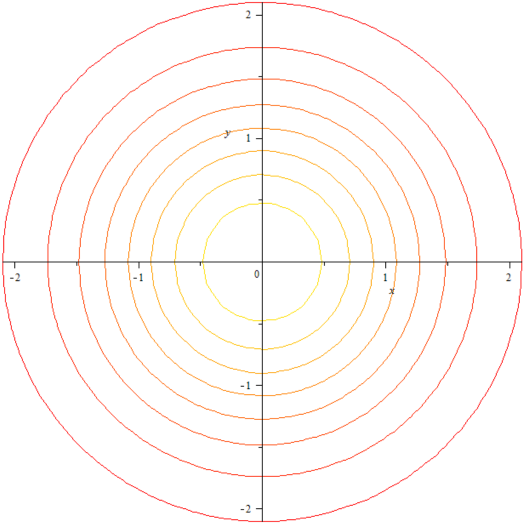
By independence, the joint p.d.f. of U and V is given by the product of the individual p.d.f.s, $f_{U,V}(u,v) = f_U(u)f_V(v)$, so $f_{U,V}(u,v)$ is

$$\frac{1}{2\pi\sigma_1\sigma_2}\exp\left\{-\frac{1}{2}\left[\frac{(u-\mu_1)^2}{\sigma_1^2}+\frac{(v-\mu_2)^2}{\sigma_2^2}\right]\right\}$$

for $(u,v)^T \in \mathbb{R}^2$.

This is a first example of a multivariate normal.

Contour plot



The general bivariate case

Let the random vector $(U, V)^T$ be defined as in the previous section, with both means zero and both variances 1, so that

$$f_{U,V}(u, v) = \frac{1}{2\pi} \exp\left\{-\frac{1}{2}\left[u^2 + v^2\right]\right\}.$$

Let

$$S = \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix}$$

be a non-singular 2×2 matrix.

Let $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ be a 2-vector.

Transformation

We now consider the random vector

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = S \begin{pmatrix} U \\ V \end{pmatrix} + \boldsymbol{\mu}.$$

We can consider this as a transformation of the random vector $\begin{pmatrix} U \\ V \end{pmatrix}$, and so we can use the theory given in section 2.4.

Forward and inverse transformations

The forward transformation is given by

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = S \begin{pmatrix} u \\ v \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}.$$

The inverse transformation is given by

$$\begin{pmatrix} u \\ v \end{pmatrix} = S^{-1} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix}.$$

Jacobian

Using the form of the inverse matrix, we can re-write the inverse transformation as

$$u = \frac{1}{\det S}\Big(s_{22}(x_1 - \mu_1) - s_{12}(x_2 - \mu_2)\Big)$$
$$v = \frac{1}{\det S}\Big(-s_{21}(x_1 - \mu_1) + s_{11}(x_2 - \mu_2)\Big).$$

The Jacobian of the inverse transformation is $|1/\det S|$.

Transformed p.d.f.

Hence the joint p.d.f. of X_1 and X_2, $f_{X_1,X_2}(x_1,x_2)$, is

$$\frac{1}{2\pi|\det S|}\exp\Big\{-\Big([s_{22}(x_1-\mu_1)-s_{12}(x_2-\mu_2)]^2+ [-s_{21}(x_1-\mu_1)+s_{11}(x_2-\mu_2)]^2\Big)/2(\det S)^2\Big\},$$

which can be rearranged as

$$\frac{1}{2\pi|\det S|}\exp\Big\{\Big(\sigma_2^2(x_1-\mu_1)^2+\sigma_1^2(x_2-\mu_2)^2 -2\sigma_{12}(x_1-\mu_1)(x_2-\mu_2)\Big)/2(\det S)^2\Big\},$$

where $\sigma_1^2=s_{11}^2+s_{12}^2$, $\sigma_2^2=s_{21}^2+s_{22}^2$ and $\sigma_{12}=s_{22}s_{12}+s_{21}s_{11}$.

Mean vector

The transformation we have used is linear, so we can also use the theory of section 2.5 to work out the mean vector and covariance matrix of $(X_1, X_2)^T$.

They are

$$E\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = S\begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

and . . .

Covariance matrix

$$
\begin{aligned}
\Sigma = \mathrm{Cov}\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} &= S \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} S^T = SS^T \\
&= \begin{pmatrix} s_{11}^2 + s_{12}^2 & s_{22}s_{12} + s_{21}s_{11} \\ s_{22}s_{12} + s_{11}s_{21} & s_{21}^2 + s_{22}^2 \end{pmatrix} \\
&= \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix},
\end{aligned}
$$

where $\sigma_1^2, \sigma_2^2, \sigma_{12}$ are defined as above.

(So σ_1^2 and σ_2^2 really are the variances of X_1 and X_2, and σ_{12} really is their covariance, as suggested by the choice of notation.)

General form

Note that $\det \Sigma = \det(SS^T) = (\det S)^2$, so we can replace $|\det S|$ by $\sqrt{\det \Sigma}$ in the above.

As $\det \Sigma = \sigma_1^2\sigma_2^2 - \sigma_{12}^2$, we can re-write the joint p.d.f. f_{X_1,X_2} as

$$\frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2 - \sigma_{12}^2}} \exp\left\{\Big(\sigma_2^2(x_1-\mu_1)^2 - 2\sigma_{12}(x_1-\mu_1)(x_2-\mu_2) + \sigma_1^2(x_2-\mu_2)^2\Big)/2(\sigma_1^2\sigma_2^2 - \sigma_{12}^2)\right\},$$

for all $\mathbf{x} = (x_1, x_2)^T \in \mathbb{R}^2$.

Notation

By analogy with the univariate case we write $\mathbf{X} \sim N_2(\boldsymbol{\mu}, \Sigma)$, where $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ and

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix},$$

the covariance matrix.

We say that the random vector $\mathbf{X}$ follows the bivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ.

Matrix form

Note that the joint p.d.f. can also be written in terms of the matrix Σ as

$$f(x_1, x_2) = \frac{1}{2\pi(\det(\Sigma))^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu})\right\}.$$

Family of bivariate normals

This p.d.f. can be defined for any symmetric positive definite 2×2 matrix Σ. (See notes for explanation.)

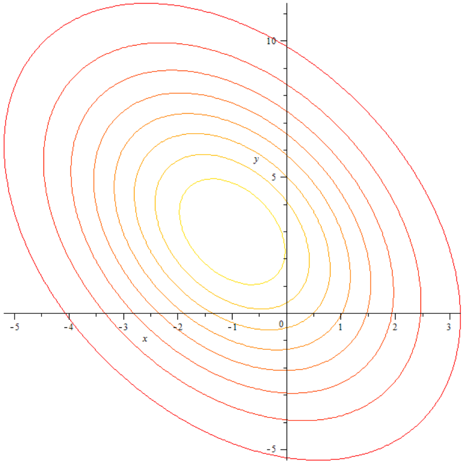
For a 2×2 symmetric matrix, being positive definite is equivalent to $\sigma_1^2, \sigma_2^2 > 0$ and $\sigma_{12}^2 < \sigma_1^2 \sigma_2^2$.

If Σ is a diagonal matrix, so $\sigma_{12} = 0$, then we recover the independent case.

Contours

The contours of f are concentric ellipses centred on $\boldsymbol{\mu}$.

Plots



$$\mu_1 = -1, \mu_2 = 3, \sigma_1 = 2, \sigma_2 = 4, \sigma_{12} = -3$$

Marginal distributions are normal

Taking marginal distributions preserves normality.

To see this, using the derivation of the bivariate normal we can write the components of a bivariate normal random vector $\mathbf{X}$ as $X_1 = s_{11}U + s_{12}V + \mu_1$ and $X_2 = s_{21}U + s_{22}V + \mu_2$ where U and V are independent standard normal random variables.

The theory of the univariate normal distribution now tells us that the marginal distributions of the components are univariate normals: $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$.

It is also possible to use the method of section 2.2.2 directly, see exercise 33(a).

Correlation, covariance, and independence

If the components of a bivariate normal have covariance (and thus correlation) zero, then they are independent.

This can be seen by letting $\sigma_{12} = 0$ in the form for the p.d.f. of a multivariate normal; the joint p.d.f. then factorises into two univariate normal p.d.f.s.

(It is important to remember that this result does **not** hold for random variables in general. It is possible to find pairs of random variables which are not independent but have correlation zero; see for example Exercise 24.)

Linear transformations of the bivariate normal

Suppose $\mathbf{X} \sim N_2(\boldsymbol{\mu}, \Sigma)$, for some known mean vector $\boldsymbol{\mu}$ and covariance matrix Σ.

For any non-singular 2×2 matrix A and for any 2×1 vector $\mathbf{b}$ define the linear transformation $\mathbf{Y} = A\mathbf{X} + \mathbf{b}$.

Normality preserved

We know that $\mathbf{X} = S\mathbf{U} + \boldsymbol{\mu}$ for some matrix S and vector $\boldsymbol{\mu}$, where $\mathbf{U} \sim N_2(\mathbf{0}, I)$.

So we can write

$$\mathbf{Y} = AS\mathbf{U} + A\boldsymbol{\mu} + \mathbf{b}.$$

Hence $\mathbf{Y}$ itself has a bivariate normal distribution with mean vector

$$\boldsymbol{\mu}_Y = E(\mathbf{Y}) = A\boldsymbol{\mu} + \mathbf{b} = AE(\mathbf{X}) + \mathbf{b}$$

and covariance matrix

$$AS(AS)^T = ASS^TA^T = A\,\mathrm{Cov}(\mathbf{X})A^T.$$

2D to 1D

We can replace A here by a row vector, $\mathbf{b} = b$ by a scalar, so giving $Y = A\mathbf{X} + b$ as a scalar.

Normality is again preserved, and the mean and the variance of Y are

$$\mu_Y = E(Y) = AE(\mathbf{X}) + b,$$
$$\sigma_Y^2 = \mathrm{Var}(Y) = A\,\mathrm{Cov}(\mathbf{X})A^T$$

giving a univariate normal distribution

$$\mathbf{Y} \sim N(\mu_Y, \sigma_Y^2).$$

Example

Example 23: Transformations of the bivariate normal

Conditional distributions are normal

Taking conditional distributions preserves normality, so that if $\mathbf{X} \sim N_2(\boldsymbol{\mu}, \Sigma)$ then the conditional distribution of X_2 given $X_1 = x_1$ is a univariate normal distribution.

In fact, conditional on $X_1 = x_1$, X_2 is normally distributed with mean

$$\mu_2 + \rho \frac{\sigma_2}{\sigma_1}(x_1 - \mu_1)$$

and variance $(1 - \rho^2)\sigma_2^2$.

(Full derivation in notes)

Conditional expectation and variance

A particular consequence of the above result is that

$$E(X_2|X_1) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(X_1 - \mu_1)$$

and

$$\mathrm{Var}(X_2|X_1) = (1 - \rho^2)\sigma_2^2.$$

The conditional expectation depends linearly on X_1 and the conditional variance does not depend upon X_1.

Example

Example 24: Conditional distributions for bivariate normal

Higher dimensions

We now generalise to higher dimensions.

Given a vector, $\boldsymbol{\mu}$, of length k and a $k \times k$ positive definite symmetric matrix Σ we can define the function

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{k/2}(\det(\Sigma))^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}.$$

It can be shown that this does indeed define a joint p.d.f. for any choice of $\boldsymbol{\mu}$ and Σ; this can be derived from the independent case by using a suitable transformation much as for the bivariate case.

Terminology

The joint distribution so specified is called the **multivariate normal distribution** $N(\boldsymbol{\mu}, \Sigma)$ or $N_k(\boldsymbol{\mu}, \Sigma)$.

It may be shown that it does indeed have mean vector $\boldsymbol{\mu}$ and covariance matrix Σ.

Transformations of the multivariate normal

Let $\mathbf{X} \sim N_k(\boldsymbol{\mu}, \Sigma)$ and consider a transformation of the form

$$\mathbf{Y} = A\mathbf{X} + \mathbf{b}$$

where A is a $m \times k$ matrix with $m \leq k$ and A is of full rank m so that $\mathbf{Y}$ has non-singular covariance matrix.

If $m = k$, so that A is a square matrix, then essentially the same argument as in section 2.6.5 for the bivariate case shows that

$$\mathbf{Y} \sim N_k(A\boldsymbol{\mu} + \mathbf{b}, A\Sigma A^T).$$

Preservation of normality extends to the case where $m < k$, where

$$\mathbf{Y} \sim N_m(A\boldsymbol{\mu} + \mathbf{b}, A\Sigma A^T).$$

Marginal distributions

In particular this property implies that all **marginal** distributions are (multivariate) normal.

For example if $k = 5$ and we take $m = 2, \mathbf{b} = 0$ and

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

then we see that the marginal joint distribution of X_1 and X_2 is multivariate normal.

As in the bivariate case, it is also possible to show this by integrating out variables.

Further properties of the multivariate normal

Conditional distributions preserve multivariate normality:

The conditional distribution of a set of the components, given values for the remaining components, will have a multivariate normal distribution.

Components are independent if and only if their covariance is zero; again note that this is a special property of the multivariate normal.

Hard to calculate

Recall that the (univariate) normal p.d.f. cannot be integrated explicitly and that normal probabilities have to be approximated numerically and tabulated or evaluated using a computer package.

Multivariate normal probabilities are even more difficult to evaluate unless the region of interest takes a special shape.

Example

Example 25: Transforming a multivariate normal